

LANGUAGE PROFICIENCY ASSESSMENT FROM PEDAGOGICAL AND STATISTICAL PERSPECTIVES

Mária Šikolová¹; Ludmila Koláčková²; Jiří Neubauer³; Marek Sedláček⁴

University of Defence,

^{1,2} Language Centre,

^{3,4} Faculty of Military Leadership, Department of Quantitative Methods,

Kounicova 65, 662 10 Brno, Czech Republic

e-mail: ¹maria.sikolova@unob.cz; ²ludmila.kolackova@unob.cz; ³jiri.neubauer@unob.cz;

⁴marek.sedlacik@unob.cz

Abstract

This exposition presents the preliminary findings of research focused on the comparison of English language proficiency self-assessments, instructor assessments and actual examination results. The examination results are from outcomes of the standardized language examination used in NATO countries (examinations in accordance with NATO STANAG 6001). The garnered data were gathered from a questionnaire administered to students enrolled in intensive, life-long language learning courses during one academic year. Statistical methods were applied to analyse the data obtained from the questionnaire to compare how accurately the students and their teachers were able to predict the actual examination corollary.

Keywords

Language assessment; Self-assessment; Standardized examination; Questionnaire; Data analysis.

Introduction

A number of experts approach issues of education and the assessment of the educational process from both pedagogical and statistical perspectives. Development of workers' skills and competencies is currently a prominent area of interest for scholars and workers in all fields [1].

The reasons behind launching this study are closely tied to the fact that the research team works for the University of Defence in Brno which is responsible for the military education in the Czech Republic; in this framework, it also guarantees the language education for military professionals in the whole Czech Armed Forces. The institution providing language education, the Language Centre, is an integral part of the University. Its mission is to supervise all aspects of language education (accredited university language courses, long-life language courses for military professionals, and standardized high-stakes language testing in accordance with NATO STANAG 6001). This study has been initiated in the academic environment of the University of Defence and it has been interwoven with other studies whose goals are not purely theoretical and academic, but their focus is on applied research in order to implement the research results into practice to improve performance of all stakeholders.

For our military personnel, attaining a designated level of language proficiency remains one of the most important prerequisites for career advancement. A recently issued regulation designates a specific level of language proficiency to be achieved by every military professional. Mere failure to reach this designated level of proficiency by the stipulated date

can have negative implications in terms of their career. To prove their language ability, they must successfully pass the standardized language examination in accordance with NATO STANAG 6001 at the required level.

All these facts make the examination in accordance with STANAG 6001 definitely a high-stakes test with all its advantages and disadvantages. On the one hand, this influence can be positive, if the test is appropriately constructed and if it results in positive washback. On the other hand, however, as Minarechova states, high-stakes tests put a lot of stress on teachers and often to such an extent that these tests sometimes resulted in problems to keep the teachers. [2, p. 95]. Moreover, often “most teaching time is devoted to preparing for the testing or doing the testing” [2, p. 92]. As mentioned above, teachers are often negatively affected by high-stakes tests. Thus, in our study we wanted to diminish the negative impact of the examination in accordance with STANAG 6001 and to let teachers participate by their estimations of students’ results at the exam.

With this in mind, we have decided to support them in their language endeavours and to innovate our approaches to both teaching and testing. As far as teaching is concerned, we are in agreement with Pavlíková, who states that language teachers should primarily concentrate on the way the language is used, and not to overestimate grammar and error correction [3, p. 72].

Regarding instruction, the content of teaching programmes has been redefined and teacher training seminars were designed and conducted, both in accordance with the exam descriptors. One of the aspects that in our opinion needed to be approached with more attention was the examination itself.

Firstly, it is usually understood as a motivational tool for learners and it is often administered in the form of a checklist consisting of several can do statements. It regularly contains also some goals and ambitions, as well as means of how to achieve them. Broadly speaking, the main intention of this kind of self-assessment is meant to be motivational.

Oscarson [4] studied self-assessment in learning writing skills in English within the environment of an upper secondary school. Apart from other things, she explored the views of both students and teachers concerning integrating self-assessment into learning – teaching process. In her thesis, she distinguished between self-assessment conducted individually and as group work focusing on writing skills; in addition, she has also looked into students’ self-assessment of both general writing ability called off-task and self-assessment of some particular writing tasks – called on-task. Besides that, she was also interested in improvement of assessment techniques [4]. To a certain extent, the goals of her study overlap with ours. While she focused on possible relationships between the individual students’ self-assessments and the teachers’ assessment expressed by correlation coefficients, our study searched for potential relationships between students’ self-assessment, teachers’ guess of proficiency level and the real results on the standardized proficiency examination. In terms of the skills under the investigation, Oscarson concentrated on writing, whereas our study refers to all four basic language skills. Her results have revealed that the correlation of the students’ self-assessment and the teacher’s grades concerning classroom writing task was $r = 0.37$ [4, p. 143]. As for the correlation of the students’ self-assessment and the teacher’s grades in the National Test of English writing test task, it was higher than in the previous one and statistically significant, $r = 0.45$ [4, p. 148].

Another author who has looked into a similar area is [5] who defines the self-assessment (he uses the expression self-appraisal interchangeably) as “a process by which students evaluate their own language competence”. Unlike Oscarson, he primarily looks into the comparison of the assessments in the area of academic listening. He has found out that there are significant

correlations between the assessments of both students and teachers. In addition, he discusses possible “pedagogical implications and applications of self-assessment”. The tools he has used were a listening test based on the English Testing Service and self-assessment questionnaire. This questionnaire contained items measuring cognitive processing skills, linguistic components, note-taking, knowledge of lecture structure, relating input to other materials, and memory and concentration [5]. What the authors considered to be inspiring is this division of the skill into the subskills and primarily Alderson’s idea that it can lead students to be “more intentional and targeted about developing their skills” [6]. Moreover, another implication worth mentioning is, as Graham emphasises, that this kind of feedback can be helpful to learners who might not fully realize their weaknesses and strengths unless they have used self-assessment [7].

Secondly, in academic settings, language tests constructed with the intention of self-assessment are often offered. Additionally, numerous checklists have been designed to enable learners to estimate their language level, the most famous of them being the CEFR (portfolio) in European context and Language Proficiency Self-Assessment for the purposes of US government service.

Nevertheless, for the conditions of our study we consider the concept of self-assessment as a subjective estimation of the result of the standardized examination in the four basic language skills compared with the same yardstick (standardized examination in accordance with NATO STANAG 6001). This estimation, in our opinion, reflects the extent to what the learners understand the language level descriptors and their fulfilment of them. Furthermore, the authors also gathered the teachers’ estimations of the results and finally compared these two data sets with the real language assessment. The results are commented on and possible implications are discussed.

1 Research Goal and Objectives

The authors have focused on several questions, beginning with a very general one: How do the respondents (students, learners) subjectively perceive English in terms of difficulty? Which of the four basic language skills do they find more difficult at different levels of proficiency? Besides, the authors’ team was interested in how well the students can predict their results at the standardized examination in four basic language skills and how well the teachers can predict them. Based on this, it is possible to speculate on the reasons behind the results.

2 Methodology

As a research tool, the authors have decided to design and use a questionnaire whose results and interpretation will be provided further on. A translation of the questionnaire is available in the attachment as well.

The authors have formulated the following hypotheses.

1. Students of individual levels of proficiency perceive the difficulty of English in the same way – see Question 1.
2. Students of individual levels of proficiency consider particular parts/ subtests of the exam to be of the same difficulty – see Question 2.
3. Teachers are able to correctly predict learners’ results at the exam in particular skills (subtests).
4. Students are able to correctly predict their results at the exam in particular skills (subtests).

3 Results

The broader and rather long-term goal of gathering the feedback from the students has been to improve the success rate of the students (military professionals) in life-long language courses at the University of Defence. We believe that familiarization with the examination is quite an important factor which may have a positive effect on both learners (increasing self-confidence, reducing test anxiety, improving self-reflection) and teachers (better choice of teaching materials and methods corresponding to required skills and knowledge). That is why the Centre has conducted a broad range of various activities to enhance the examination awareness of both teachers and students. As for the teachers, they have had an opportunity to attend seminars dealing with the examination in general, calibration seminars, and specialized test familiarization seminars. As far as the students are concerned, they are regularly given a presentation combined with a discussion concerning the examination and its details (length, format, way of assessment, ...). Moreover, the curriculum also comprises various tests on regular basis. However, the mock exam, which copies the format of the real exam, is of utmost importance.

The evaluation of the questionnaires has been conducted separately for three different courses, depending on their level, i.e. courses preparing for level 1 (survival), level 2 (functional), level 3 (professional). The statistical environment R was used for data analyses.

Questions to Study

Question 1: How difficult is English for you?

The first question asked the respondents how difficult English was for them, which they indicated on a scale ranging from very easy (1) to very difficult (10).

Question 2: Which of the four basic skills do you consider to be the most difficult for you?

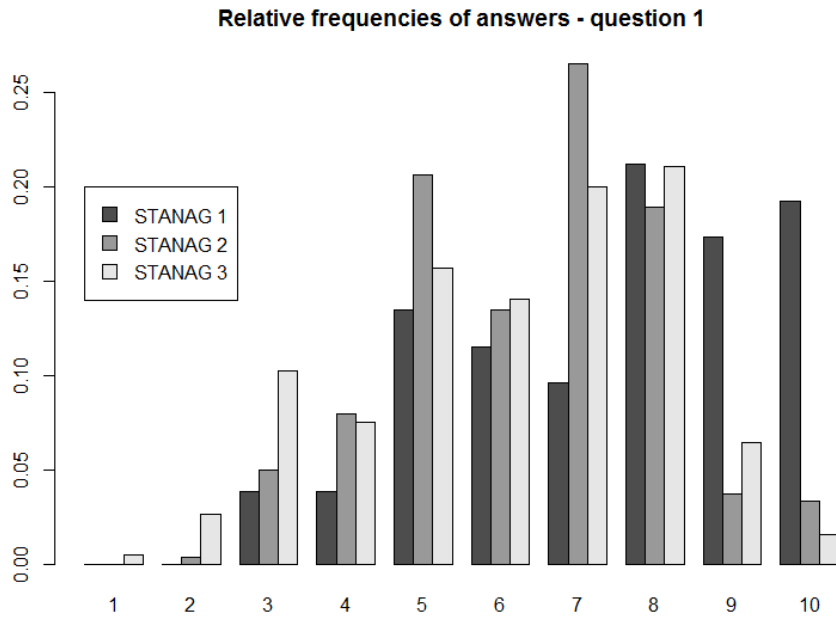
The respondents were asked to choose one of the four basic skills – listening, speaking, reading, and writing.

Question 3: What are the relationships (correlations) between the self-assessment, teachers' guess of the examination results and real examination results (everything in all four skills)?

3.1 Level 1 Courses

3.1.1 Question 1

The results in the courses aimed at achieving level 1 (survival) have shown that the students of these beginners' courses ($n = 52$) generally perceive English as quite a difficult language. Only 2 (3.8%) of the respondents find English quite easy (on the scale point 3 out of 10). Most respondents have chosen point 8 (21.2%), which is very close to point 10 (very difficult); this point on the scale was chosen by 10 respondents (19.2%), see Figure 1.

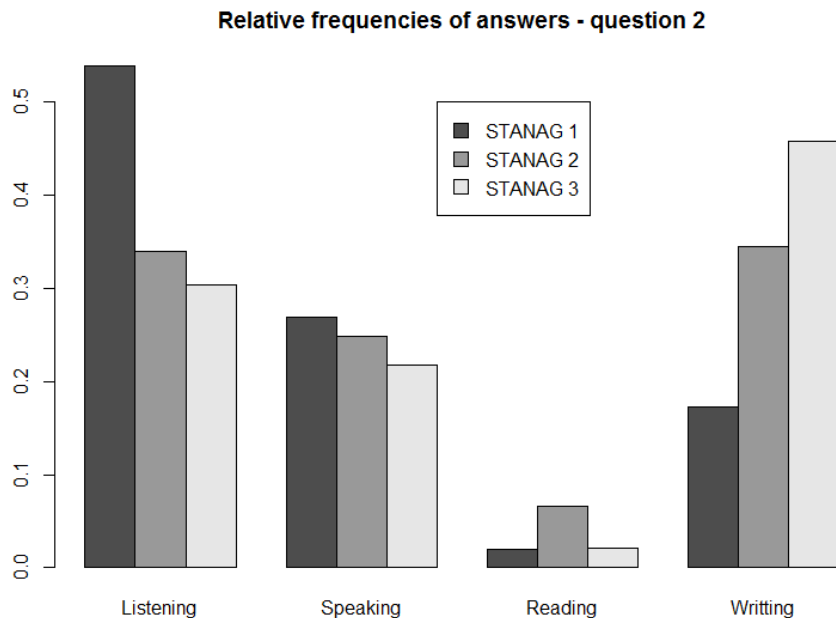


Source: Own

Fig. 1: Relative frequencies of answers to Question 1: How difficult is English for you? (very easy = 1, ..., very difficult = 10, level 1 courses)

3.1.2 Question 2

28 (54%) out of 52 respondents consider listening skill to be the most difficult one, while reading comprehension is perceived as the most difficult one only by 2% (1 respondent only), see Figure 2.



Source: Own

Fig. 2: Relative frequencies of answers to Question 2: Which of the four basic language skills do you consider to be the most difficult one for you?

3.1.3 Question 3

The results have shown that correlations between two estimations (teachers' and students') and the real results are very similar, ranging from 0.448 to 0.667 and they are statistically significant (at the significance level 0.05, see Table 1). The highest correlation of 0.667 has occurred between students' and teachers' estimation of the assessments of writing, while the lowest correlation has been found between students' estimation and the real results, also in writing. To define significant differences between the real results and the estimations by both students and teachers, we have chosen two types of *t*-test, namely parametric *t*-test and non-parametric Wilcoxon test [8, 9].

Tab. 1: Spearman's correlation coefficients of results and teachers' and students' predictions – level 1

	Listening	Speaking	Reading	Writing
Student – Result	0.522	0.508	0.544	0.448
Teacher – Result	0.540	0.500	0.544	0.541
Student – Teacher	0.546	0.552	0.546	0.667

Source: Own

Table 2 summarizes basic descriptive characteristics of the data collected.

Tab. 2: Descriptive measures of level 1 results and students' and teachers' predictions, $n = 52$

Listening	Minimum	Maximum	Mean	Median	St. dev.
Student	0	2	0.760	1	0.614
Teacher	0	2	0.856	1	0.621
Result	0	2	0.750	1	0.645
Speaking					
Student	0	2	0.750	1	0.573
Teacher	0	2	0.904	1	0.552
Result	0	2	0.827	1	0.609
Reading					
Student	0	2	0.971	1	0.572
Teacher	0	2	0.894	1	0.644
Result	0	2	0.913	1	0.677
Writing					
Student	0	2	0.692	1	0.620
Teacher	0	2	0.904	1	0.552
Result	0	2	0.721	1	0.606

Source: Own

Based on the results given in Table 3, it can be concluded that there are statistically significant differences between students' and teachers' predictions and real results in the writing skill. The results in this skill are overestimated by teachers, if compared with the real results. Similarly, the comparison of the students' and teachers' predictions shows that the teachers again predict better results even than students themselves. A couple of possible explanations come into play – either the students' performance is worse at the examination on the grounds of test anxiety, or the teachers have not fully identified and/ or calibrated with NATO STANAG 6001 descriptors, or the teachers are just not able to objectively assess the students' skills because they have known them for a longer time and might have labelled them imprecisely. As for speaking, the teachers have predicted higher results than the students,

however, no significant difference has been proven between either students' predictions and real results, or teachers' predictions and real results. As far as the estimation of the level of the other skills, no differences have been revealed.

Tab. 3: Paired *t*-test and Wilcoxon test; statistically significant differences ($\alpha = 0.05$) are in bold – level 1

Paired <i>t</i> -tests						
	Student-result	<i>p</i> -value	Teacher-result	<i>p</i> -value	Student-teacher	<i>p</i> -value
Listening	0.110	0.913	1.228	0.225	-1.121	0.268
Speaking	-0.942	0.351	0.955	0.344	-2.028	0.048
Reading	0.676	0.502	-0.234	0.816	0.929	0.357
Writing	-0.335	0.739	2.428	0.019	-2.894	0.006

Wilcoxon paired test						
	Student-result	<i>p</i> -value	Teacher-result	<i>p</i> -value	Student-teacher	<i>p</i> -value
Listening	208.5	0.634	242.0	0.189	144.5	0.418
Speaking	105.5	0.491	111.0	0.268	73.5	0.080
Reading	285.0	0.27	166.5	0.825	158.5	0.294
Writing	109.5	0.843	150.5	0.024	45.5	0.011

Source: Own

The percentage of the correct students' predictions of the assessment in Table 4 ranges from 42.3% to 59.6%, while the correct teachers' predictions are in the range of 48.1% to 65.4%. This suggests that the teachers are better aware of the real proficiency level of students than they themselves. As far as the overestimation of the real results is concerned, the highest percentage has been found in teachers when 30.8% of teachers have overestimated the results in listening comprehension test and 28.8% of them in writing skills. With the exception of the reading comprehension test, the percentage of the overestimation of teachers is always higher than that of students. Thus, it can be concluded that the teachers' expectations are higher than the students' ones. As for the underestimation of the real results, the percentage for teachers ranges from 7.7% in writing to 28.8% in reading; the percentage for students ranges from 23.1% in writing to 30.8% in reading.

Tab. 4: Ratio of level 1 result estimates

	Underestimation	Correct	Overestimation
Listening – student	0.288	0.481	0.231
Listening – teacher	0.212	0.481	0.308
Speaking – student	0.269	0.577	0.154
Speaking – teacher	0.115	0.654	0.231
Reading – student	0.308	0.423	0.269
Reading – teacher	0.288	0.500	0.212
Writing – student	0.231	0.596	0.173
Writing – teacher	0.077	0.635	0.288

Source: Own

3.2 Level 2 Courses

The results gathered from the respondents attending level 2 courses were different from those collected from level 1 courses. Also, the number of respondents was much higher ($n = 238$).

3.2.1 Question 1

In level 2 courses, the most frequently used was point 7 (26.5%), while in level 1 courses it was point 8 (21.2%). English perceived as a very difficult language (point 10 out of 10) was in level 1 courses by 19.2% of the respondents, while in level 2 courses only 3.4% of the respondents, see Figure 1.

3.2.2 Question 2

The students of level 2 courses consider both listening (34%) and writing (34%) to be the most difficult skills – out of 241 respondents 83 have chosen writing and 82 listening. If compared with the results in level 1 courses, the listening is again perceived as the most difficult one, however, not as difficult as in level 1 courses – 34% in level 2 in comparison with 54% in level 1 courses. Reading skill was in both kinds of courses evaluated as the most difficult one by the least number of respondents – only 16 (7%) in level 2 courses and 1 respondent (2%) in level 1 courses, see Figure 2.

3.2.3 Question 3

The existence of statistical link between individual estimations and the achieved level of proficiency can be found out by means of correlation coefficients, see Table 5. All of them are positive and statistically significant. Therefore, there is a dependency between individual assessments. If compared with the correlation coefficients in level 1 courses, these coefficients are lower which implies that in level 2 courses, the relation between the estimations and real results is weaker. The strongest correlation has been revealed between students' and teachers' estimations in listening skills (0.543), while the lowest one occurred in the estimations of students and real results in reading skills (0.212).

Tab. 5: Spearman's correlation coefficients of results and teachers' and students' predictions – level 2

	Listening	Speaking	Reading	Writing
Student – Reality	0.331	0.267	0.212	0.389
Teacher – Reality	0.366	0.393	0.316	0.471
Student – Teacher	0.543	0.425	0.454	0.466

Source: Own

The results in Tables 6 and 7 have shown that in speaking skills, there is an obvious difference between students' estimations and the real result. A similar difference has been identified in teachers' estimations. Both students and teachers have overestimated the real performance. In addition, other rather different assessments have occurred in listening comprehension where teachers' estimations are, in average, higher than students' estimations. In writing skills, teachers' estimates were too optimistic.

Table 8, which illustrates ratios of underestimations, correct predictions and overestimations, shows that a very similar percentage of both students (34.3%) and teachers (34.6%) were too optimistic in their estimations of the results in speaking skills. The highest ratio of correct estimations (61.2%) was in reading comprehension by teachers.

Tab. 6: Descriptive measures of level 2 results and students' and teachers' predictions, $n = 52$

	Listening	Minimum	Maximum	Mean	Median	St. dev.
Student	0	0	2.5	1.690	2	0.399
Teacher	0	0	2.5	1.751	2	0.360
Result	0	0	2.5	1.723	2	0.395
Speaking						
Student	0	0	2.5	1.677	1.75	0.376
Teacher	1	0	2.5	1.716	2	0.361
Result	0	0	2.5	1.625	1.5	0.376
Reading						
Student	0	0	3	1.767	2	0.368
Teacher	0	0	3	1.797	2	0.320
Result	1	0	3	1.783	2	0.315
Writing						
Student	0	0	2.5	1.569	1.5	0.399
Teacher	0	0	3	1.618	1.5	0.405
Result	1	0	2.5	1.565	1.5	0.359

Source: Own

Tab. 7: Paired t -test and Wilcoxon test; statistically significant differences ($\alpha = 0.05$) are in bold – level 2

Paired t -tests

	Student-result	p -value	Teacher-result	p -value	Student-teacher	p -value
Listening	-1.004	0.316	1.141	0.255	-2.563	0.011
Speaking	1.987	0.048	3.636	0	-1.483	0.139
Reading	-0.525	0.600	0.611	0.542	-1.220	0.224
Writing	0.232	0.817	2.230	0.027	-1.845	0.066

Wilcoxon paired test

Listening	2865.0	0.210	2882.0	0.266	1165	0.012
Speaking	5958.0	0.048	4992.0	0	2463	0.146
Reading	3005.0	0.617	2346.5	0.500	1678	0.276
Writing	3901.5	0.807	3609.0	0.025	2442	0.068

Source: Own

Tab. 8: Ratio of level 2 result – estimates

	Underestimation	Correct	Overestimation
Listening – student	0.264	0.523	0.213
Listening – teacher	0.188	0.579	0.233
Speaking – student	0.251	0.406	0.343
Speaking – teacher	0.162	0.492	0.346
Reading – student	0.238	0.531	0.230
Reading – teacher	0.175	0.612	0.212
Writing – student	0.234	0.485	0.280
Writing – teacher	0.158	0.550	0.292

Source: Own

3.3 Level 3 Courses

3.3.1 Question 1

Interestingly, in level 3 courses ($n = 185$), most of the respondents consider English to be fairly difficult – point 8 was chosen by most of the respondents (21.1%). As for English perceived as rather easy (point 3), 10.3 % of respondents have chosen it in comparison with only 3.8 % in level 1 courses and 5% in level 2 courses. As for perceiving English as very difficult (point 10), the differences are rather big, oscillating between 19.2% in level 1 courses, 3.4% in level 2 courses and 1.6% in level 3 courses, see Figure 1.

3.3.2 Question 2

In level 3 courses the skill which was chosen as the most difficult one by most respondents was writing – 86 respondents out of 188 (46%). Listening was considered as the most difficult one by 57 respondents (30%), while speaking by 41 respondents (22%). Reading skill was considered as the most difficult one only by 4 respondents (2%) which is similar to the results in both L1 (2%) and L2 courses (7%), see Figure 2.

3.3.3 Question 3

The analysis of the results of level 3 courses has been conducted analogically as the one of the results from level 1 and 2 courses. Correlation coefficients in Table 9 are all statistically significant due to the size of the data set, although some values are relatively low, for example the correlation coefficient between the students' estimates and the actual achieved level in speaking is only 0.229. The highest correlation value of 0.511 was found in the pair of students' and teachers' predictions of listening skills.

Tab. 9: Spearman's correlation coefficients of results and teachers' and students' predictions – level 3

	Listening	Speaking	Reading	Writing
Student – Reality	0.437	0.229	0.235	0.265
Teacher – Reality	0.462	0.380	0.322	0.507
Student – Teacher	0.511	0.401	0.410	0.327

Source: Own

Based on the descriptive characteristics in Table 10 and the paired tests in Table 11 it can be stated that in both speaking and writing skills both students and teachers statistically significantly overestimated the level actually achieved. These conclusions correspond to the underestimation and overestimation rates shown in Table 12 where, for example, 44.4% of students and 42% of teachers overestimated the actual level in writing. The percentage of overestimation was slightly lower for speaking, 36% for students and 38.7% for teachers. In terms of other skills, these percentages are significantly lower.

Tab. 10: Descriptive measures of level 3 results and student's and teacher's predictions, $n = 188$

	Listening	Minimum	Maximum	Mean	Median	St. dev.
Student	2	2	3	2.709	3	0.331
Teacher	2	2	3	2.718	3	0.323
Result	2	2	3	2.726	3	0.329
Speaking						
Student	2	2	3	2.649	2.5	0.349
Teacher	1.5	1.5	3	2.694	3	0.363
Result	2	2	3	2.543	2.5	0.369
Reading						
Student	2	2	3	2.816	3	0.282
Teacher	2	2	3	2.809	3	0.302
Result	2	2	3	2.822	3	0.266
Writing						
Student	1.5	1.5	3	2.556	2.5	0.371
Teacher	1.5	1.5	3	2.574	2.5	0.383
Result	2	2	3	2.395	2.5	0.391

Source: Own

Tab. 11: Paired t -test and Wilcoxon test, statistically significant differences ($\alpha = 0.05$) are in bold – level 3

Paired t -tests

	Student-result	p -value	Teacher-result	p -value	Student-teacher	p -value
Listening	-0.625	0.533	-0.108	0.914	-0.441	0.660
Speaking	3.362	0.001	5.589	0	-1.606	0.110
Reading	-0.218	0.828	-0.555	0.58	0.337	0.737
Writing	5.183	0	6.661	0	-0.670	0.503

Wilcoxon paired test

	Student-result	p -value	Teacher-result	p -value	Student-teacher	p -value
Listening	1281.5	0.528	1024	0.909	978.5	0.652
Speaking	3195	0.002	3558	0	1315	0.111
Reading	1276.5	0.816	1123.5	0.576	992.5	0.712
Writing	4712	0	3680	0	1833	0.457

Source: Own

Tab. 12: Ratio of level 3 result estimates

	Underestimation	Correct	Overestimation
Listening – student	0.211	0.600	0.189
Listening – teacher	0.172	0.656	0.172
Speaking – student	0.161	0.478	0.360
Speaking – teacher	0.124	0.489	0.387
Reading – student	0.186	0.617	0.197
Reading – teacher	0.191	0.633	0.176
Writing – student	0.178	0.378	0.444
Writing – teacher	0.099	0.481	0.420

Source: Own

3.4 Comparison of the Answers to Question 1

The comparison of the answers to Question 1 according to individual language proficiency levels in accordance with STANAG 6001 is presented in Figure 1. Basic descriptive characteristics are summarized in Table 13.

Tab. 13: Descriptive statistics – Question 1

Question 1	<i>n</i>	Minimum	Maximum	Mean	Median	St. dev.
Level 1	52	3	10	7.481	8	2.024
Level 2	238	2	10	6.357	7	1.682
Level 3	185	1	10	6.130	6	1.960

Source: Own

Based on the results of Pearson χ^2 test of independence (p-value is equal to $1.62 \cdot 10^{-7}$), it can be stated that the answers to the Question 1 are dependent on the level of English proficiency. Both Figure 7 and Table 13 show that the students on level 1 consider English to be more difficult than students on level 3. Tests in Table 14 show that for the students on level 1, English is more difficult than for more proficient students, i.e. for the students on levels 2 and 3. As for the comparison of the answers of the students on levels 2 and 3, no significant difference has been observed.

Tab. 14: Two-sample *t*-test a Wilcoxon test

	<i>t</i> -test	<i>p</i> -value	Wilcoxon	<i>p</i> -value
level 1 – level 2	4.200	$3.56 \cdot 10^{-5}$	8212	$8.26 \cdot 10^{-5}$
level 1 – level 3	4.360	$1.95 \cdot 10^{-5}$	6592.5	$3.68 \cdot 10^{-5}$
level 2 – level 3	1.283	0.200	22923	0.460

Source: Own

3.5 Comparison of the Answers to Question 2

The comparison of the answers of the students of different levels in accordance with STANAG 6001 to Question 2 is graphically presented in Figure 1.

Similarly as in Question 1, the dependence of the answers to the questions on the proficiency level has been verified by Pearson χ^2 test of independence. It can be stated that the answers to Question 2 are dependent on the proficiency level in English (p-value is equal to 0.001). Table 15 contains relative frequencies of the answers to Question 2 which show that for the students on level 1, the most difficult language skill is listening, while the easiest one is reading. For the students on level 2, both listening and writing are the most difficult ones and reading is the easiest one. However, for the students on level 3, there is a significant grow in the difficulty in writing skill.

Tab. 15: Relative frequencies – Question 2

	Listening	Speaking	Reading	Writing
Level 1	0.538	0.269	0.019	0.173
Level 2	0.340	0.249	0.066	0.344
Level 3	0.303	0.218	0.021	0.457

Source: Own

4 Discussion

The hypothesis that students of individual levels of proficiency perceive the difficulty of English in the same way has been disproved. The difference in the assessment of difficulty of English has been proved between students on level 1 and levels 2 and 3. Nevertheless, the

difference in the assessment of the difficulty of English has not been detected between students on levels 2 and 3. Students on level 1 perceive English to be more difficult than the students on higher levels (2 and 3). The reasons behind this, in our opinion, are in the fact that for adults, to learn the basics of the language which they have never been exposed to could be very difficult. Once they have acquired at least the pre-intermediate level and they build on the basic skills and knowledge, the language is probably perceived as less difficult.

The hypothesis that students of all levels consider the subtests of each skill to be on the same level of difficulty has also been disproved. For the students on level 1, listening is the most difficult out of four basic language skills. This could be because they have not mastered the phonetical system of English and hence, they have not got used to recognize and process spoken texts. For the students on level 2, apart from listening also writing is considered as the most difficult skill. This could be connected with the fact that the writing tasks on level 2 are getting more complex than on level 1 and they require the higher level of overall language proficiency. As for the students on level 3, the perception of the difficulty of writing has grown significantly.

The hypothesis that teachers are able to correctly predict the students' results at the exam in individual skills has been partially disproved. In writing skills, teachers overestimated the results of students of all levels. In speaking skills, they significantly overestimated the results of the students on levels 2 and 3. This could have been caused by either the exam anxiety of students resulting in worse performance than regularly shown in lessons, or by some discrepancies in the interpretation of the exam descriptors by teachers.

The hypothesis that the students are able to correctly predict their exam results in particular skills has also been partially disproved. For level 1 students no difference between the predicted and real results has been proven. Nevertheless, the students on level 2 have overestimated their results in speaking skills and the students on level 3 have overestimated their results not only in speaking, but also in writing skills. From our perspective, we believe that the overestimation of students' results by themselves is interconnected with the similar approach of the teachers who work in the closest contact with students and therefore they presumably interpret the descriptors to students in the same way, as they interpret them by themselves. Another possible factor coming into play could again be exam anxiety.

For the last two hypotheses it can be concluded that there is a mismatch in the perception of level 2 and 3 productive skills by both teachers and students and the real required language level.

Conclusion

As far as the students' perception of the difficulty of English is concerned, it varies depending on the level for which the courses prepare (1–3), see Figure 1. The higher the level of English proficiency, the lower the perception of the difficulty of English is. The more proficient the learners are, the less difficult they perceive the English language. This is reflected in average difficulty of English from level 1 (the mean 7.48 out of 10, the median 8), through level 2 (the mean 6.36, the median 7) up to level 3 (the mean 6.13, the median 6).

As far as the most difficult skill is concerned, for level 1 respondents it was represented by listening, while level 2 respondents added writing, which became the most difficult for level 3 respondents.

Surprisingly, teachers tend to overestimate students' results, nevertheless, this overestimation in terms of skills depends on the level. Thus in level 1 courses the teachers overestimate mostly the results in writing, in level 2 courses it is in speaking, and in level 3 courses both

speaking and writing have been overestimated by teachers. The reasons might be various; either the teachers are not sufficiently familiarized and/ or identified with NATO STANAG 6001 descriptors, or the learners' performance can be affected by test anxiety and hence, it is lower at the examination than in the classroom; or the teachers might just not be objective enough, labelling the students according to a complex impression. Another fact that is worth mentioning is that all teachers' overestimations are in productive, not receptive skills, although it is apparently productive skills which are directly displayed, since the level of receptive skills is more hidden and hence, more difficult to measure. Moreover, although the skill of writing in level 3 courses is perceived as the most difficult one, still the expectations are overoptimistic, both for students and teachers. This could be again caused by insufficient familiarization with what is required at this level by teachers and learners alike.

For future application of the results, we suggest continuing in closer cooperation between teachers and testers in terms of common seminars with the emphasis on both tests familiarization and calibration. Another contribution which we should focus on in the near future is raising test awareness of students at particular proficiency levels and skills by means of continuous feedback of their progress resulting in their higher ability of self-reflection.

Acknowledgements

This research was supported by LANGUAGE (DZRO CJV) project.

Literature

- [1] ČECHOVÁ, I.; SALIGER, R.: Development of Academic Staff' Competencies. In: *Innovation Management and Education Excellence Vision 2020: from Regional Development Sustainability to Global Economic Growth, Vols I – VI*. Italy: International Business Information Management Association (IBIMA). 2016, pp. 1900–1906.
- [2] MINARECHOVÁ, M.: Negative impacts of high-stakes testing. *Journal of Pedagogy*. 2012, Vol 3, Issue 1, pp. 82–100. DOI: [10.2478/v10159-012-0004-x](https://doi.org/10.2478/v10159-012-0004-x)
- [3] PAVLÍKOVÁ, S.: Profesní rozvoj učitelů angličtiny ve 21. století. *ACC JOURNAL*. 2017, Vol. 23, Issue 3, pp. 70–78. DOI: [10.15240/tul/004/2017-3-006](https://doi.org/10.15240/tul/004/2017-3-006)
- [4] OSCARSON, A. D.: *Self-Assessment of Writing in Learning English as a Foreign Language; A Study at the Upper Secondary School Level*. Göteborgs: Universitet, Göteborg, 2009.
- [5] ARYADOUST, V.: *Reliability of Second Language Listening Self-Assessments: Implications for Pedagogy*. Singapore: National University of Singapore, 2012.
- [6] ALDERSON, J. Ch.: Diagnosing foreign language proficiency: A teaching/testing interface. In: *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. pp. 4–12. London: Bloomsbury Academic, 2005. DOI: [10.5040/9781474212151.ch-001](https://doi.org/10.5040/9781474212151.ch-001)
- [7] GRAHAM, S.: Listening comprehension: The learners' perspective. *System*. 2006, Vol. 34, Issue 2, pp. 165–182. DOI: [10.1016/j.system.2005.11.001](https://doi.org/10.1016/j.system.2005.11.001)
- [8] DEVORE, D. L.: *Probability and Statistics for Engineering and the Sciences*. Brooks/Cole, Cengage Learning, 2012.
- [9] MONTGOMERY, D. C.; RUNGER, G. C. *Applied Statistics and Probability for Engineers*. 5th ed. John Wiley & Sons, 2011.

PhDr. Mária Šikolová, Ph.D.; Mgr. Ludmila Koláčková, Ph.D.;
doc. Mgr. Jiří Neubauer, Ph.D.; doc. RNDr. Marek Sedlačík, Ph.D.

HODNOCENÍ JAZYKOVÉ KOMPETENCE Z PERSPEKTIVY PEDAGOGIKY A STATISTIKY

Článek pojednává o výsledcích studie založené na srovnání, jak svoji jazykovou kompetenci hodnotí sami studenti, jejich učitelé a jaké jsou skutečné výsledky dosažené u zkoušky. Jedná se o výsledky standardizované jazykové zkoušky, kterou používají členské státy NATO (zkouška podle STANAG 6001). Prezentovaná data byla získána z dotazníku distribuovaného studentům intenzivních jazykových kurzů v rámci celoživotního vzdělávání během jednoho akademického roku. Pomocí aplikovaných statistických metod byla analyzována data z dotazníků, aby bylo možno srovnat, jak přesně jsou studenti a jejich učitelé schopni předvídat výsledky zkoušek.

BEURTEILUNG DER SPRACHKENNTNISSE AUS PÄDAGOGISCHER UND STATISTISCHER SICHT

Der Artikel behandelt die vorläufigen Ergebnisse einer Studie zum Vergleich der Selbstbewertung der englischen Sprachkenntnisse, der Beurteilung der Lehrer und der tatsächlichen Testergebnisse. Die Testergebnisse stammen aus standardisierten Sprachtests (NATO-STANAG-6001-Tests), die in den NATO-Ländern verwendet werden. Die präsentierten Daten stammen aus einem Fragebogen, der während eines akademischen Jahres von den Studenten ausgefüllt wurde, die Intensivkurse im Rahmen des lebenslangen Lernens besucht haben. Zur Datenanalyse des Fragebogens wurden statistische Methoden angewendet, um zu vergleichen, wie genau die Studenten und ihre Lehrer die Testergebnisse vorhersagen können.

OCENA KOMPETENCJI JĘZYKOWYCH Z PERSPEKTYWY PEDAGOGIKI I STATYSTYKI

Niniejszy artykuł poświęcony jest wynikom badań dotyczących porównania samooceny kompetencji językowej dokonanej przez studentów oraz oceny dokonywanej przez nauczycieli z faktycznymi wynikami osiągniętymi na egzaminie. Wykorzystano w tym celu wyniki zestandaryzowanego egzaminu z języka, stosowanego w krajach członkowskich NATO (egzamin wg STANAG 6001). Zaprezentowane dane zebrano w drodze ankiety przeprowadzonej wśród studentów intensywnych kursów językowych w ramach kształcenia ustawicznego w ciągu jednego roku akademickiego. Przy pomocy stosowanych metod statystycznych dane z ankiet przeanalizowano w celu dokonania porównania, na ile dokładnie studenci i ich nauczyciele są w stanie przewidzieć wyniki egzaminów.